

Maschinelles Lernen

Übungsblatt 8

Dr. Steffen Rendle, Christoph Freudenthaler
Wirtschaftsinformatik und Maschinelles Lernen (ISMLL)
Universität Hildesheim

12. Januar 2010
Abgabe bis 21. Januar 2010

Support Vektor Maschinen I - Lineare Separierung, Margin

Nr.	Zielklasse	X_1	X_2
D1	-	1	2
D2	-	2	3
D3	+	4	2

a) Lineare Separierung Wiederholen Sie den Begriff der linearen Separierung. Was bedeutet er? Bezieht er sich auf die Daten, die Modellklasse oder ein Lösungsverfahren der Support-Vektor-Maschinen? Sind die vorliegenden Daten linear separierbar? Warum? Erstellen Sie ein Streudiagramm mit den obigen Datenpunkten.

b) Margin Definieren Sie mit eigenen Worten den Begriff des Margins. Nehmen Sie an, dass $\beta = (1, 1)$ und $\beta_0 = -5$. Zeichnen Sie die so definierte Trenngerade in das Streudiagramm ein. Wie hoch ist die Missklassifikationsrate dieser Trenngerade. Sehen Sie eine bessere Möglichkeit zur Separierung der positiven von der negativen Zielklasse? Zeichnen Sie ihren Vorschlag ein und geben Sie für ihren Vorschlag und $\beta = 1, 1, \beta_0 = -5$ den Margin an. Warum ist es sinnvoll eine Trenngerade mit maximalem Margin zu verlangen?

Support Vektor Maschinen II - Perceptron

a) Führen Sie für die vorliegenden Daten den Perceptron-Algorithmus durch. Zeichnen Sie die resultierende Trenngerade in das Streudiagramm ein. Verwenden Sie als Startwerte für den Perceptron-Algorithmus $\beta = (0, 0), \beta_0 = 0$.

b) Der Perceptron-Algorithmus kann auf zwei Arten berechnet werden: Vollständiger Gradientenabstieg und stochastischer Gradientenabstieg. Inwieweit unterscheiden sich beide Verfahren? Erklären Sie anhand eines Beispiels.

c) Berechnen Sie den Margin der neuen Trenngerade. Definieren Sie in eigenen Worten das Gütekriterium einer Trenngeraden (oder allgemein Hyperebene) mit maximalem Margin. Welche der drei Trenngeraden ist die beste gemessen an diesem Gütekriterium.

Support Vektor Maschinen III

Führen Sie für die vorliegenden Daten den Lernalgorithmus für SVMs durch und merken sie sich für jeden Iterationsschritt die ermittelten Werte für α .

b) Berechnen Sie pro Iterationsschritt β_0 , β_1 und β_2 . Zeichnen Sie weiters die resultierenden Trenngeraden pro Iterationsschritt in das Streudiagramm ein und kennzeichnen Sie die Sequenz der gefundenen Trenngeraden. Berechnen Sie pro Zwischenschritt den Margin und beobachten Sie den Verlauf dieser Größe. Verwenden Sie als Startwerte für den SVM-Algorithmus $\alpha = (1, 1, 1)$. Bestimmen Sie abschließend welche der Trenngeraden, die Sie für diesen Datensatz ermittelt haben, ist die beste, gemessen an diesem Gütekriterium.

c) Was ist ein Supportvektor? Benennen Sie die Supportvektoren in Ihrer Lösung.